

Application of Local Rank Tests to Nonparametric Regression

Lutz Dümbgen
University of Berne

August 1998, revised December 1999

Appeared in:

Journal of Nonparametric Statistics **14** (2002), 511–537

Abstract. Let $Y_i = f(x_i) + E_i$ ($1 \leq i \leq n$) with given covariates $x_1 < x_2 < \dots < x_n$, an unknown regression function f and independent random errors E_i with median zero. It is shown how to apply several linear rank test statistics simultaneously in order to test monotonicity of f in various regions and to identify its local extrema.

Keywords and phrases. exponential inequality, linear rank statistic, modality, monotonicity, multiscale testing, quadratic complexity

1 Introduction

Suppose that one observes $(x_1, Y_1), (x_2, Y_2), \dots, (x_n, Y_n)$, where $x_1 < x_2 < \dots < x_n$ are given real numbers, and the Y_i are independent random variables with continuous distribution functions $G_i(\cdot) := \mathbb{P}\{Y_i \leq \cdot\}$. With $\mathbf{G} := ((x_i, G_i))_{1 \leq i \leq n}$ we call \mathbf{G} increasing on an interval $J \subset \mathbb{R}$ if

$$G_i \leq_{\text{st.}} G_j \quad \text{whenever } x_i, x_j \in J \text{ and } x_i \leq x_j.$$

Here $G_i \leq_{\text{st.}} G_j$ means that G_i is stochastically smaller than G_j , that means, $G_i \geq G_j$ pointwise. Analogously we call \mathbf{G} decreasing on J if $G_i \geq_{\text{st.}} G_j$ for $x_i, x_j \in J$ with $x_i \leq x_j$.

A special case of this setup is the standard nonparametric regression model, where

$$(1.1) \quad Y_i = f(x_i) + E_i$$

with an unknown regression function f and independent random errors E_i having continuous distribution function G and mean or median zero. Then $G_i(\cdot) = G(\cdot - f(x_i))$, whence \mathbf{G} is increasing or decreasing on J , provided that f is monotonically increasing or decreasing on J , respectively.

Our goal is to identify intervals J on which \mathbf{G} is very unlikely to be decreasing or increasing. In other words, we aim at finding intervals on which a monotone trend in the data is significant. This is similar to Chaudhuri and Marron's (1998) approach. A major difference is that instead of linear kernel estimators we use *local rank tests*. Because of this modification we don't need any further distributional assumptions. Another difference is that in the model (1.1) we always interpret our results in terms of the function f , whereas Chaudhuri and Marron focus on the convolution of f with a Gaussian kernel with varying bandwidth.

Here is the definition of the local rank tests: For integers $0 \leq j < k \leq n$ let $R_{jk}(i)$ be the rank of Y_i among the $k - j$ numbers Y_{j+1}, \dots, Y_k , i.e.

$$R_{jk}(i) := \sum_{h=j+1}^k 1\{Y_h \leq Y_i\}.$$

A monotone trend of the latter observations is quantified via the linear rank statistic

$$T_{jk} := \sum_{i=j+1}^k \beta\left(\frac{i-j}{k-j+1}\right) q\left(\frac{R_{jk}(i)}{k-j+1}\right),$$

where β, q are functions on $]0, 1[$ such that

$$\beta(1-u) = -\beta(u) \text{ and } q(1-u) = -q(u) \text{ for } 0 < u < 1,$$

$$\beta \text{ and } q \text{ are non-decreasing on }]0, 1[.$$

For later purposes we define $\beta(x) := q(x) := 0$ for $x \in \mathbb{R} \setminus]0, 1[$. All these local rank statistics T_{jk} are considered simultaneously. We define the multiscale test statistic

$$T := \max_{0 \leq j < k \leq n} (s_{k-j}|T_{jk}| - c_{k-j})$$

with arbitrary nonnegative numbers s_d, c_d . Specific suggestions for these normalizing constants are given later. Here is a key observation:

Theorem 1.1 *Let T_{jk}^* and T^* be defined as T_{jk} and T , respectively, where the local rank $R_{jk}(i)$ is replaced with*

$$R_{jk}^*(i) := \sum_{h=j+1}^k 1\{U_h \leq U_i\},$$

and U_1, U_2, \dots, U_n are independent random variables with uniform distribution on $[0, 1]$. Then for arbitrary $r \in \mathbb{R}$, the probability of the event

$$\begin{aligned} & \bigcup_{0 \leq j < k \leq n} \left\{ s_{k-j}T_{jk} > c_{k-j} + r \text{ and } \mathbf{G} \text{ is decreasing on } [x_{j+1}, x_k] \right\} \\ & \cup \bigcup_{0 \leq j < k \leq n} \left\{ -s_{k-j}T_{jk} > c_{k-j} + r \text{ and } \mathbf{G} \text{ is increasing on } [x_{j+1}, x_k] \right\} \end{aligned}$$

is not greater than $\mathbb{P}\{T^ > r\}$.*

In other words, let $\kappa(\alpha)$ be the $(1 - \alpha)$ -quantile of the distribution of T^* and define

$$\begin{aligned} \mathcal{D}^+(\alpha) &:= \left\{ [x_{j+1}, x_k] : 0 \leq j < k \leq n, s_{k-j}T_{jk} > c_{k-j} + \kappa(\alpha) \right\}, \\ \mathcal{D}^-(\alpha) &:= \left\{ [x_{j+1}, x_k] : 0 \leq j < k \leq n, -s_{k-j}T_{jk} > c_{k-j} + \kappa(\alpha) \right\}. \end{aligned}$$

Then with probability at least $1 - \alpha$ the following conclusion is correct: \mathbf{G} is neither decreasing on any interval in $\mathcal{D}^+(\alpha)$ nor increasing on any interval in $\mathcal{D}^-(\alpha)$.

Let us consider the special model (1.1) with a continuous regression function f . The local rank tests imply confidence bounds for the number and location of local extrema of f . Namely, let

$$\begin{aligned} \mathcal{M}^+(\alpha) &:= \left\{ [x, \tilde{y}] : [x, y] \in \mathcal{D}^+(\alpha), [\tilde{x}, \tilde{y}] \in \mathcal{D}^-(\alpha), y \leq \tilde{x} \right\}, \\ \mathcal{M}^-(\alpha) &:= \left\{ [x, \tilde{y}] : [x, y] \in \mathcal{D}^-(\alpha), [\tilde{x}, \tilde{y}] \in \mathcal{D}^+(\alpha), y \leq \tilde{x} \right\}, \\ \mathcal{M}(\alpha) &:= \left\{ [x \wedge \tilde{x}, y \wedge \tilde{y}] : [x, y] \in \mathcal{D}^-(\alpha), [\tilde{x}, \tilde{y}] \in \mathcal{D}^+(\alpha) \right\} \supset \mathcal{M}^+(\alpha) \cup \mathcal{M}^-(\alpha). \end{aligned}$$

Then with probability at least $1 - \alpha$ any interval J in $\mathcal{M}(\alpha)$ contains points $u < v < w$ such that $f(v) \notin [\min\{f(u), f(w)\}, \max\{f(u), f(w)\}]$. More precisely,

$$f(v) \begin{cases} > \max\{f(u), f(w)\} & \text{if } J \in \mathcal{M}^+(\alpha), \\ < \min\{f(u), f(w)\} & \text{if } J \in \mathcal{M}^-(\alpha). \end{cases}$$

This procedure implies a lower $(1 - \alpha)$ -confidence bound for the number of local extrema of f . As shown by Donoho (1988), a nontrivial upper confidence bound for this number cannot exist.

The remainder of this paper is organized as follows. Section 2 proposes special constants s_d, c_d and provides asymptotic properties of T^* . Section 3 gives some results on the asymptotic power of our procedure in case of the Wilcoxon score function, $q(u) = u - 1/2$, and the standard model (1.1). An algorithm for the computation of $(T_{jk})_{j,k}$ in $O(n^2)$ steps in case of $q(u) = \beta(u) = 2u - 1$ is provided in Section 4. Explicit computer code in ‘MatLab’ is available from the author via the internet. Finally, we illustrate our methods in Section 5. Some useful inequalities for linear rank statistics, some of which are new and potentially of independent interest, are compiled in Section 6. All proofs are deferred to Section 7.

Remark on ties. For the sake of simplicity we assume pairwise different values x_i and continuous distribution functions G_i . In general, if merely $x_1 \leq x_2 \leq \dots \leq x_n$, one can proceed as follows:

(i) Replace $\beta((i - j)/(k - j + 1))$ with the average of $\beta((h - j)/(k - j + 1))$ over all indices h such that $x_h = x_i$.

(ii) Redefine

$$R_{jk}(i) := \sum_{h=j+1}^k \left(1\{Y_h < Y_i\} + 1\{Y_h = Y_i\}/2 \right) + 1/2.$$

(iii) Consider T_{jk} only for pairs (j, k) such that $x_j < x_{j+1}$ (or $j = 0$) and $x_k < x_{k+1}$ (or $k = n$). One can show that this leads to confidence sets with asymptotic confidence level at least $1 - \alpha$ as n tends to infinity.

2 Normalizing constants

So far we have not specified the constants $s_d, c_d \geq 0$. It is well-known from the theory of linear rank tests that in case of $k - j > 1$,

$$\text{Var}(T_{jk}^*) = \frac{\Sigma_{k-j}(\beta)^2 \Sigma_{k-j}(q)^2}{k - j - 1},$$

where

$$\Sigma_d(g) := \left(\sum_{i=1}^d g\left(\frac{i}{d+1}\right)^2 \right)^{1/2}$$

for any function g on the real line. For $g = \beta, q$ the number $d^{-1} \Sigma_d(g)^2$ is close to $\sigma(g)^2$, where

$$\sigma(g) := \left(\int_0^1 g(x)^2 dx \right)^{1/2}.$$

Thus we define

$$s_d := \frac{1 + h_d}{\sigma(\beta)\sigma(q) d^{1/2}}$$

with arbitrary constants h_d such that $|h_d| = O(d^{-1})$. As for the additive correction terms c_d , recent results of Dümbgen and Spokoiny (1999) about multiscale testing in gaussian white noise models suggest

$$c_d := C(d/n) \quad \text{with} \quad C(r) := \left(2 \log(1/r)\right)^{1/2};$$

see also Proposition 7.1.

In order to reduce the computational burden and to increase sensitivity on smaller intervals one may restrict one's attention to pairs (j, k) with $k - j \leq m$ for some integer m in $[1, n]$. Thus we define generally

$$T := \max_{1 \leq d \leq m} T_d \quad \text{with} \quad T_d := \max_{j, k \in \{0, 1, \dots, n\} : k-j=d} s_d |T_{jk}| - c_d.$$

The test statistics T^* and T_d^* are defined in the same way with uniform random variables in place of the Y_i . In the definition of $\mathcal{D}^\pm(\alpha)$ the index pairs (j, k) are restricted accordingly.

The test statistic T^* has a nondegenerate limiting distribution as n tends to infinity, provided that the functions β and q satisfy the following additional requirements:

$$(2.1) \quad \beta \text{ is bounded with } \sigma(\beta) > 0;$$

$$(2.2) \quad q \text{ is convex on } [1/2, 1[\text{ with } 0 < \sigma(q) < \infty;$$

$$(2.3) \quad M(q, \lambda) := \int_0^1 \exp(\lambda q(x)) dx \leq \exp\left(\frac{\sigma(q)^2 \lambda^2}{2}\right) \quad \text{for all } \lambda \in \mathbb{R};$$

$$(2.4) \quad \sup_{0 < \delta \leq 1} \delta^{-1} \int_0^{1-\delta} \left(q(u + \delta) - q(u)\right)^2 du < \infty.$$

Theorem 2.1 *Suppose that Conditions (2.1–2.4) are satisfied. Let T^* be defined with $m = \lfloor un \rfloor$ for some fixed $0 < u \leq 1$. Then it converges in distribution, as $n \rightarrow \infty$, to the random variable*

$$T^o := \sup_{s, t \in [0, 1] : 0 < t-s \leq u} \left(\frac{1}{\sigma(\beta)(t-s)^{1/2}} \left| \int_0^1 \beta\left(\frac{x-s}{t-s}\right) dW(x) \right| - C(t-s) \right),$$

where W is a standard Brownian motion on $[0, 1]$. Moreover, $0 \leq T^o < \infty$ almost surely.

Remark 1. Theorem 2.1 does not require monotonicity of β . The conclusion remains true if β has just bounded total variation. As for the stochastic integral in the definition of T^o , the present assumptions about β imply the existence of a finite signed measure μ on the unit interval such that $\mu([0, 1]) = 0$ and $\mu([0, x]) = \beta(x)$ for all but at most countably many $x \in [0, 1]$. Then it follows from partial integration that

$$\int_0^1 \beta\left(\frac{x-s}{t-s}\right) W(dx) = - \int_{[0, 1]} W(s + (t-s)y) \mu(dy).$$

Remark 2. Conditions (2.2–2.4) on the score function q are satisfied in the following three cases:

Wilcoxon score function. Let $q(x) = 2x - 1$. Thus q is the quantile function of the uniform distribution on $[-1, 1]$ with variance $\sigma(q)^2 = 1/3$. Its moment generating function equals $M(q, \lambda) = \sinh(\lambda)/\lambda \leq \exp(\lambda^2/6)$.

“Triangular” score function. Let $q(x) = 1\{x > 1/2\}(1 - (2 - 2x)^{1/2}) - 1\{x < 1/2\}(1 - (2x)^{1/2})$. This is the quantile function of the distribution with triangular density $f(x) = (1 - |x|)^+$, variance $\sigma(q)^2 = 1/6$ and moment generating function $M(q, \lambda) = 2\lambda^{-2}(\cosh(\lambda) - 1) \leq \exp(\lambda^2/12)$.

Van der Waerden score function. Let $q = \Phi^{-1}$, the standard Gaussian quantile function. Here $\sigma(q)^2$ equals one, and $M(q, \lambda) = \exp(\lambda^2/2)$.

The inequalities involving \sinh and \cosh can be deduced from the corresponding series expansion. The Wilcoxon and the triangular score function satisfy Condition (2.4) because they are Hölder-continuous with exponent 1 and $1/2$, respectively. The proof of (2.4) for the van der Waerden score function is more involved and deferred to Section 7.

3 Asymptotic power

We study the performance of our procedure in case of the Wilcoxon score function $q(u) = 2u - 1$ and the standard regression model (1.1) with equidistant design points $x_i = (i - 1/2)/n$. As for the error distribution function G , we assume that

$$(3.1) \quad \liminf_{\delta \rightarrow 0} \delta^{-1} \left(\int G(x + \delta) G(dx) - 1/2 \right) > 0.$$

This condition is satisfied whenever G admits a Lebesgue density.

Theorem 3.1 *Suppose that for some constants $x_o \in]0, 1[$ and $\gamma \in]0, 2]$,*

$$(3.2) \quad \lim_{x \rightarrow x_o} \frac{f(x_o) - f(x)}{|x_o - x|^\gamma} = \lambda \neq 0.$$

Further let f be monotone on $[x_o - \epsilon, x_o]$ and on $[x_o, x_o + \epsilon]$ for some $\epsilon > 0$. Then, as $n \rightarrow \infty$,

$$\inf \left\{ y - x : x_o \in [x, y] \in \mathcal{M}^{\text{sign}(\lambda)}(\alpha) \right\} = O_p \left(\left(\frac{\log n}{n} \right)^{1/(2\gamma+1)} \right),$$

where $\inf(\emptyset) := \infty$.

Note that Theorem 3.1 holds for arbitrary fixed levels $\alpha \in]0, 1[$. Thus we detect any local extremal point x_o of f satisfying (3.2) with asymptotic probability one, and we can estimate x_o with precision $O_p \left((\log(n)/n)^{1/(2\gamma+1)} \right)$. In case of $\gamma = 2$ we end up with the rate $O_p \left((\log(n)/n)^{1/5} \right)$. This is close to the familiar optimal rate $O_p((1/n)^{1/5})$ for estimating the mode of a density; cf. Khas'minski (1979).

4 Computational aspects

Since T is a maximum of $O(n^2)$ single random variables its computation is necessarily expensive. But even devising an algorithm requiring $O(n^2)$ steps (quadratic complexity) is far from trivial. A naive algorithm would compute the local ranks $R_{jk}(\cdot)$ first, using some advanced sorting routine requiring $O((k-j) \log(k-j))$ steps. Then the computation of T_{jk} requires another $O(k-j)$ steps. Since we are dealing with $O(n^2)$ such pairs, we would end up with $O(n^3 \log n)$ steps.

The problem of the naive approach is that it does not utilize the fact that many of our index intervals $]j, k]$ overlap. When sorting and ranking values Y_i in a certain interval of indices, this information should be utilized somehow for larger intervals. Although this is a natural demand, it is not obvious how to fulfill it.

Now we treat the special case of the Wilcoxon score function q and $\beta = q$. For this case we present an algorithm avoiding the computation of ranks at all. It computes T_{jk} in $O((k-j)^2)$ steps, which seems to be disastrous at first glance. But the computation of all local rank statistics together requires $O(n^2)$ steps as well. Note first that by Proposition 6.7,

$$T_{jk} = (k-j+1)^{-2} \sum_{j < a < b \leq k} 2(b-a) \operatorname{sign}(Y_b - Y_a) = 2(k-j+1)^{-2} \tilde{T}_{jk},$$

where

$$\tilde{T}_{jk} := \sum_{j < a < b \leq k} (b-a) \operatorname{sign}(Y_b - Y_a).$$

Given these values \tilde{T}_{jk} , the computation of

$$\begin{aligned} T &= \max_{(j,k)} (\tilde{s}_{k-j} |\tilde{T}_{jk}| - c_{k-j}), \\ \mathcal{D}^\pm(\alpha) &= \left\{ (j,k) : \pm \tilde{s}_{k-j} \tilde{T}_{jk} > c_{k-j} + \kappa(\alpha) \right\} \end{aligned}$$

requires $O(n^2)$ steps, where $\tilde{s}_d := 2s_d(d+1)^{-2}$ and, for instance, $s_d := \Sigma_d(q)^{-2}(d-1)^{1/2} = 3(1+1/d)(d-1)^{-1/2}$. Thus we only need an algorithm with quadratic complexity for the computation of all numbers \tilde{T}_{jk} together. Basic recursion formulae are:

$$\begin{aligned} \tilde{T}_{jk} &= \tilde{T}_{j,k-1} + S_{jk} \quad \text{with } S_{jk} := \sum_{j < \ell < k} (k-\ell) \operatorname{sign}(Y_k - Y_\ell), \\ S_{jk} &= S_{j+1,k} + (k-j-1) \operatorname{sign}(Y_k - Y_{j+1}). \end{aligned}$$

Here is a suitable algorithm utilizing these recursions. One easily verifies that it does return all numbers \tilde{T}_{jk} in $O(n^2)$ steps:

ALGORITHM $Y \mapsto (\tilde{T}_{jk})_{0 \leq j < k \leq n}$

$$\tilde{T}_{01} \leftarrow 0$$

```

FOR  $k \leftarrow 2$  TO  $n$  DO
   $S \leftarrow 0$ 
   $\tilde{T}_{k-1,k} \leftarrow 0$ 
  FOR  $j \leftarrow k-2$  DOWNTO  $0$  DO
     $S \leftarrow S + (k-j-1) \text{sign}(Y_k - Y_{j+1})$ 
     $\tilde{T}_{jk} \leftarrow \tilde{T}_{j,k-1} + S$ 
  END
END

```

This implementation is easy to understand but would require storing all $O(n^2)$ numbers \tilde{T}_{jk} . Representing and storing the set $\mathcal{D}^\pm(\alpha)$ is an additional problem. We propose to replace $\mathcal{D}^\pm(\alpha)$ with the set $\mathbf{D}^\pm(\alpha)$ of its *minimal* elements. An interval J in $\mathcal{D}^\pm(\alpha)$ is called ‘minimal’ if $\mathcal{D}^\pm(\alpha)$ contains no proper subset of J . The set $\mathbf{D}^\pm(\alpha)$ contains at most $n-1$ intervals because their left endpoints are pairwise different and contained in $\{x_1, x_2, \dots, x_{n-1}\}$. Here is an algorithm for the computation of T and $\mathbf{D}^\pm(\alpha)$ in $O(mn)$ steps. It utilizes two vector variables $T^{(1)}, T^{(2)} \in \mathbb{R}^m$ and an integer variable h^\pm with the following purpose: For $k = 2, 3, \dots, n$,

$$\begin{aligned}
T^{(1)}(d) &= \tilde{T}_{k-1-d,k-1} \quad (1 \leq d \leq \min(k-1, m)), \\
T^{(2)}(d) &= \tilde{T}_{k-d,k} \quad (1 \leq d \leq \min(k, m)), \\
h^\pm &= \max\left(\{0\} \cup \left\{a : [a, b] \in \mathbf{D}^\pm(\alpha), b \leq x_k\right\}\right).
\end{aligned}$$

ALGORITHM $Y \mapsto (T, \mathbf{D}^\pm(\alpha))$

```

 $T \leftarrow -c_1$ 
 $\mathbf{D}^\pm(\alpha) \leftarrow \emptyset$ 
 $T^{(1)} \leftarrow 0$ 
 $T^{(2)} \leftarrow 0$ 
 $h^\pm \leftarrow 0$ 
FOR  $k \leftarrow 2$  TO  $n$  DO
   $S \leftarrow 0$ 
  FOR  $d \leftarrow 2$  TO  $\min(k, m)$  DO
     $S \leftarrow S + (d-1) \text{sign}(Y_k - Y_{k-d+1})$ 
     $T^{(2)}(d) \leftarrow T^{(1)}(d-1) + S$ 
     $T \leftarrow \max\left(T, \tilde{s}_d |T^{(2)}(d)| - c_d\right)$ 
    IF  $(k-d+1 > h^\pm)$  AND  $(\pm \tilde{s}_d T^{(2)}(d) > c_d + \kappa(\alpha))$  THEN
       $\mathbf{D}^\pm(\alpha) \leftarrow \left\{ \mathbf{D}^\pm(\alpha), [x_{k-d+1}, x_k] \right\}$ 

```


$$h^\pm \leftarrow k - d + 1$$

END

END

$$T^{(1)} \leftarrow T^{(2)}$$

END

Given $D^\pm(\alpha)$, one can easily compute the minimal elements of $\mathcal{M}^\pm(\alpha)$ as well as the minimal elements of $\mathcal{M}(\alpha)$ with $O(n)$ steps.

5 Numerical examples

Again we consider the special case $q(u) = \beta(u) = 2u - 1$ and let $n = 500$. At first we demonstrate the distribution of T^* . Figure 1 shows three realizations of the stochastic processes $d \mapsto T_d^*$ as well as the additive correction function $d \mapsto -c_d$. Apparently the normalizing constants s_d and c_d work well.

Figure 1 around here

Now we consider T and T^* with $m = \lfloor n/2 \rfloor$. We estimated the distribution function of T^* in 9999 Monte-Carlo simulations; see Figure 2. The vertical lines highlight selected estimated quantiles: $\hat{\kappa}(0.5) = 1.519$, $\hat{\kappa}(0.1) = 2.191$ and $\hat{\kappa}(0.05) = 2.401$.

Figure 2 around here

Figure 3 depicts simulated data (x_i, Y_i) , $1 \leq i \leq n$, from the standard model (1.1) with errors E_i having logistic distribution. The underlying regression function, shown as gray line, has two local maxima and one local minimum in $]0, 1[$.

Figure 3 around here

The corresponding stochastic process $d \mapsto T_d$ is shown in Figure 4. Its maximum clearly exceeds the critical value $\hat{\kappa}(0.05)$ which is indicated by a horizontal line.

Figure 4 around here

Figures 5a and 5b show minimal intervals in $\mathcal{D}^+(0.05)$ and $\mathcal{D}^-(0.05)$, respectively. Each such interval $J = [x_{j+1}, x_k]$ is represented as a horizontal line segment. Both families $\mathcal{D}^\pm(0.05)$ reflect the fact that the underlying regression function is bimodal. Note also the different lengths of the intervals.

Figure 5a around here

Figure 5b around here

The resulting family $\mathcal{M}^+(0.05)$ contains ten minimal intervals:

$$\begin{aligned} &[0.229, 0.403], [0.233, 0.407], [0.245, 0.761], [0.393, 0.769], [0.435, 0.771], \\ &[0.451, 0.773], [0.459, 0.775], [0.477, 0.777], [0.479, 0.779], [0.483, 0.781]. \end{aligned}$$

The family $\mathcal{M}^-(0.05)$ contains one minimal interval, namely $[0.295, 0.553]$. Since there are two non-overlapping intervals in $\mathcal{M}^+(0.05)$ while $\mathcal{M}^-(0.05)$ is nonvoid, one can conclude with 95 percent confidence that the regression function f has at least two local maxima and at least one local minimum in $]0, 1[$.

6 Auxiliary results for linear rank statistics

Throughout this section let $Y_i = G_i^{-1}(U_i)$ with the quantile function $G_i^{-1}(u) := \min\{r \in \mathbb{R} : G_i(r) \geq u\}$ of G_i . By continuity of G_i , its quantile function is strictly increasing. In particular, if we ignore events with probability zero, then $R(i) := R_{0n}(i) = \sum_{j=1}^n 1\{Y_j \leq Y_i\}$ and $R^*(i) := R_{0n}^*(i) = \sum_{j=1}^n 1\{U_j \leq U_i\}$ define random variables R, R^* with values in the set \mathcal{S}_n all permutations of $\{1, 2, \dots, n\}$. The random permutation R^* is uniformly distributed on \mathcal{S}_n . Here we consider linear rank statistics

$$\sum_{i=1}^n b_i a_{R(i)},$$

where a, b are some vectors in \mathbb{R}^n . For notational simplicity we write $v_+ := \sum_{i=1}^n v_i$, $\|v\|_2 := \left(\sum_{i=1}^n v_i^2\right)^{1/2}$ and $\|v\|_\infty := \max_{1 \leq i \leq n} |v_i|$ for vectors $v \in \mathbb{R}^n$.

In order to formulate our first inequality let us define a partial order relation on \mathcal{S}_n . For $1 \leq k < \ell \leq n$ and $\pi \in \mathcal{S}_n$ define $E_{k\ell}\pi \in \mathcal{S}_n$ via

$$E_{k\ell}\pi(i) := \begin{cases} \pi(i) & \text{if } i \notin \{k, \ell\}, \\ \pi(\ell) & \text{if } i = k, \\ \pi(k) & \text{if } i = \ell. \end{cases}$$

For $\pi, \rho \in \mathcal{S}_n$ we write $\pi < \rho$ if there exist finitely many permutations $\rho_0 = \pi, \rho_1, \dots, \rho_m = \rho$ in \mathcal{S}_n such that $\rho_z = E_{k(z), \ell(z)}\rho_{z-1}$ for $1 \leq z \leq m$, where $k(z) < \ell(z)$ and $\rho_{z-1}(k(z)) > \rho_{z-1}(\ell(z))$. The following two inequalities are the key to Theorem 1.1.

Proposition 6.1 *Suppose that $G_1 \leq_{\text{st.}} G_2 \leq_{\text{st.}} \dots \leq_{\text{st.}} G_n$. Then $R^* \leq R$.*

Proposition 6.2 *Let $\pi, \rho \in \mathcal{S}_n$ with $\pi \leq \rho$. Then*

$$\sum_{i=1}^n \beta\left(\frac{i}{n+1}\right) q\left(\frac{\pi(i)}{n+1}\right) \leq \sum_{i=1}^n \beta\left(\frac{i}{n+1}\right) q\left(\frac{\rho(i)}{n+1}\right).$$

The subsequent inequalities are required for the proof of Theorem 2.1.

Proposition 6.3 *Let $a, b \in \mathbb{R}^n$ such that $a_+ = 0$ or $b_+ = 0$. Then for arbitrary $\eta \geq 0$,*

$$\mathbb{P}\left\{\sum_{i=1}^n b_i a_{R^*(i)} > \eta\right\} \leq \exp\left(-\frac{\eta^2}{128\|b\|_\infty^2\|a\|_2^2}\right).$$

Proposition 6.4 *Let $a, b \in \mathbb{R}^n$ such that $a_+ = 0$. Then*

$$\mathbb{E}\left(\sum_{i=1}^n b_i a_{R^*(i)} - \sum_{i=1}^n b_i a_{\lceil nU_i \rceil}\right)^2 \leq 6n^{-1}\|a\|_2^2\|b\|_2\|b\|_\infty.$$

Proposition 6.3 is essentially Corollary 1 (a) of Dümbgen (1998), while Proposition 6.4 is due to Hájek (1961).

Proposition 6.5 *Suppose that the function q satisfies (2.2). Then $\Sigma_n(q)^2/n \leq \sigma(q)^2$ and for arbitrary $b \in \mathbb{R}^n$,*

$$\mathbb{E} \exp\left(\sum_{i=1}^n b_i q\left(\frac{R^*(i)}{n+1}\right)\right) \leq \mathbb{E} \exp\left(\sum_{i=1}^n b_i q(U_i)\right).$$

Proposition 6.5 implies exponential inequalities for linear rank statistics with Wilcoxon, triangular or van der Waerden scores:

Corollary 6.6 *Suppose that the function q satisfies (2.2–2.3). Then for arbitrary vectors $b \in \mathbb{R}^n$ and $\eta \geq 0$,*

$$\begin{aligned} \text{Var}\left(\sum_{i=1}^n b_i q\left(\frac{R^*(i)}{n+1}\right)\right) &\leq \sigma(q)^2\|b\|_2^2, \\ \mathbb{P}\left\{\pm \sum_{i=1}^n b_i q\left(\frac{R^*(i)}{n+1}\right) \geq \eta\right\} &\leq \exp\left(-\frac{\eta^2}{2\sigma(q)^2\|b\|_2^2}\right). \end{aligned}$$

For the proof of the results in Section 3 we need an elementary bound for linear rank statistics using Wilcoxon scores.

Proposition 6.7 *For any $b \in \mathbb{R}^n$,*

$$\begin{aligned} \sum_{i=1}^n b_i \left(2\frac{R(i)}{n+1} - 1\right) &= \frac{1}{n+1} \sum_{1 \leq i < j \leq n} (b_j - b_i) \text{sign}(Y_j - Y_i) \\ &= \frac{1}{n+1} \sum_{1 \leq i < j \leq n} (b_j - b_i) \left(2 \int G_i dG_j - 1\right) + Z, \end{aligned}$$

where

$$\mathbb{E} Z^2 \leq \|b\|_2^2 \left(1 + 12 \max_{1 \leq i, j \leq n} \|G_i - G_j\|_\infty^2\right).$$

7 Proofs

Proof of Theorem 1.1. As in Section 6 let $Y_i = G_i^{-1}(U_i)$. It follows from Propositions 6.1 and 6.2, applied to $\{j+1, \dots, k\}$ in place of $\{1, 2, \dots, n\}$, that $T_{jk} \geq T_{jk}^*$ if \mathbf{G} is increasing on $[x_{j+1}, x_k]$. Symmetry considerations show that $T_{jk} \leq T_{jk}^*$ if \mathbf{G} is decreasing on $[x_{j+1}, x_k]$. Consequently, T^* is an upper bound for the maximum of

$$\max_{(j,k)} 1 \left\{ \mathbf{G} \text{ is decreasing on } [x_{j+1}, x_k] \right\} (s_{k-j}T_{jk} - c_{k-j})$$

and

$$\max_{(j,k)} 1 \left\{ \mathbf{G} \text{ is increasing on } [x_{j+1}, x_k] \right\} (-s_{k-j}T_{jk} - c_{k-j}). \quad \square$$

Proof of Proposition 6.1. For $1 \leq \ell \leq n$ let

$$Y_{\ell,i} := \begin{cases} G_i^{-1}(U_i) & \text{if } i \leq \ell, \\ G_\ell^{-1}(U_i) & \text{if } i \geq \ell, \end{cases}$$

$$R_\ell(i) := \sum_{j=1}^n 1\{Y_{\ell,j} \leq Y_{\ell,i}\}.$$

Then $R_\ell \in \mathcal{S}_n$ almost surely, and $R_1 = R^*$, $R_n = R$. Thus it suffices to show that $R_{\ell-1} \leq R_\ell$ for any fixed $\ell \in \{2, \dots, n\}$.

Note that $Y_{\ell-1,i} = Y_{\ell,i}$ for $1 \leq i < \ell$, that $Y_{\ell-1,i} \leq Y_{\ell,i}$ for $\ell \leq i \leq n$, and that the tuple $(Y_{\ell-1,i})_{\ell \leq i \leq n}$ has the same order as $(Y_{\ell,i})_{\ell \leq i \leq n}$. Thus

$$(7.1) \quad (R_{\ell-1}(i))_{1 \leq i < \ell} \text{ and } (R_\ell(i))_{1 \leq i < \ell} \text{ have the same order,}$$

$$(7.2) \quad R_{\ell-1}(i) \leq R_\ell(i) \text{ for } \ell \leq i \leq n,$$

$$(7.3) \quad (R_{\ell-1}(i))_{\ell \leq i \leq n} \text{ and } (R_\ell(i))_{\ell \leq i \leq n} \text{ have the same order.}$$

If $R_{\ell-1} \neq R_\ell$, let k be the index such that $R_\ell(k)$ equals $\max\{R_\ell(i) : i \geq \ell, R_{\ell-1}(i) < R_\ell(i)\}$. Let j be the index such that $R_{\ell-1}(j) = R_{\ell-1}(k) + 1$. This index j is smaller than ℓ . For if $j \geq \ell$ it would follow from (7.3) that

$$R_\ell(j) > R_\ell(k) \geq R_{\ell-1}(k) + 1 = R_{\ell-1}(j),$$

a contradiction to the definition of k . Conditions (7.1–7.3) remain valid with $E_{jk}R_{\ell-1}$ in place of $R_{\ell-1}$. Moreover, $R_{\ell-1} < E_{jk}R_{\ell-1}$ and

$$(7.4) \quad \sum_{i=\ell}^n (R_\ell(i) - E_{jk}R_{\ell-1}(i)) = \sum_{i=\ell}^n (R_\ell(i) - R_{\ell-1}(i)) - 1.$$

If the lefthand side of (7.4) equals zero, then $E_{jk}R_{\ell-1} = R_\ell$. Otherwise one may replace $R_{\ell-1}$ with $E_{jk}R_{\ell-1}$ and repeat the preceding considerations. After finitely many steps we arrive at R_ℓ , whence $R_{\ell-1} \leq R_\ell$. \square

Proof of Proposition 6.2. It suffices to consider the case $\rho = E_{jk}\pi$ with $j < k$ and $\pi(j) > \pi(k)$. Then

$$\begin{aligned} & \sum_{i=1}^n \beta\left(\frac{i}{n+1}\right) q\left(\frac{\rho(i)}{n+1}\right) - \sum_{i=1}^n \beta\left(\frac{i}{n+1}\right) q\left(\frac{\pi(i)}{n+1}\right) \\ &= \left(\beta\left(\frac{k}{n+1}\right) - \beta\left(\frac{j}{n+1}\right)\right) \left(q\left(\frac{\pi(j)}{n+1}\right) - q\left(\frac{\pi(k)}{n+1}\right)\right) \geq 0, \end{aligned}$$

by isotonicity of β and q . \square

Proof of Condition (2.4) for van der Waerden scores. Here q is the standard normal quantile function. It follows from the smoothness of q and symmetry considerations that (2.4) is equivalent to

$$\sup_{0 < \delta \leq 1/4} \delta^{-1} \int_0^{1/4} \left(q(u + \delta) - q(u)\right)^2 du < \infty.$$

Denoting the standard normal density and distribution function with ϕ and Φ , respectively, it is wellknown that $\Phi(x)$ equals $|x|^{-1}\phi(x)(1 + o(1))$ as $x \rightarrow -\infty$. This implies that, as $u \rightarrow 0$,

$$q(u) = -\left(2 \log(1/u) - \log \log(1/u) - \log(4\pi) + o(1)\right)^{1/2},$$

whence

$$\begin{aligned} \frac{dq(u)}{du} = \phi(q(u))^{-1} &= (2\pi)^{1/2} \exp(\log(1/u) - (\log \log(1/u))/2 - (\log(4\pi))/2 + o(1)) \\ &= u^{-1} (2 \log(1/u))^{-1/2} (1 + o(1)). \end{aligned}$$

Since $\tilde{q}(u) := -(2 \log(1/u))^{1/2}$ has derivative $u^{-1} (2 \log(1/u))^{-1/2}$, assertion (2.4) is equivalent to

$$\sup_{0 < \delta \leq 1/4} \delta^{-1} \int_0^{1/4} \left(\tilde{q}(u + \delta) - \tilde{q}(u)\right)^2 du < \infty.$$

But $\tilde{q} \geq 2^{1/2}$ on $]0, 1/4]$, whence

$$\begin{aligned} \delta^{-1} \int_0^{1/4} \left(\tilde{q}(u + \delta) - \tilde{q}(u)\right)^2 du &= \delta^{-1} \int_0^{1/4} \left(\frac{\tilde{q}(u + \delta)^2 - \tilde{q}(u)^2}{\tilde{q}(u + \delta) + \tilde{q}(u)}\right)^2 du \\ &\leq 2\delta^{-1} \int_0^{1/4} (\log(1 + \delta/u))^2 du \\ &\leq 2 \int_0^\infty (\log(1 + 1/x))^2 dx \\ &< \infty. \quad \square \end{aligned}$$

Here is a key ingredient for the proof of Theorem 2.1. It is a consequence of Theorem 6.1 of Dümbgen and Spokoiny (1999).

Proposition 7.1 *Let Y be a stochastic process on a subset Π of $\{(s, t) : 0 \leq s < t \leq 1\}$ having continuous sample paths. Suppose that the following inequalities hold for real constants $L, M > 0$ and arbitrary $(s, t), (s', t') \in \Pi, \eta \geq 0$:*

$$(7.5) \quad \mathbb{P}\left\{|Y(s, t)| > (t - s)^{1/2}\eta\right\} \leq L \exp(-\eta^2/2),$$

$$(7.6) \quad \mathbb{P}\left\{|Y(s, t) - Y(s', t')| > \rho((s, t), (s', t'))\eta\right\} \leq M \exp(-\eta^2/M),$$

where $\rho((s, t), (s', t')) := (|s - s'| + |t - t'|)^{1/2}$. Then the random variable

$$S(Y) := \sup_{(s, t) \in \Pi} \left(\frac{|Y(s, t)|}{\sqrt{t - s}} - C(t - s) \right) / D(t - s)$$

is finite almost surely, where $D(h) := (\log(e/h))^{-1/2} \log \log(e/h)$. More precisely, there is a universal function p depending only on the constants L, M such that $\lim_{\eta \rightarrow \infty} p(\eta) = 0$ and $\mathbb{P}\{S(Y) > \eta\} \leq p(\eta)$. \square

Proof of Theorem 2.1. Without loss of generality one may assume that $\sigma(q) = \sigma(\beta) = 1$. At first one may apply Proposition 7.1 to the Gaussian process Y_o on $\Pi_o := \{(s, t) \in [0, 1]^2 : 0 < t - s \leq u\}$ with

$$(7.7) \quad Y_o(s, t) := \int_0^1 \beta\left(\frac{x - s}{t - s}\right) W(dx);$$

cf. Dümbgen and Spokoiny (1999, proof of Theorem 2.1). Since T_o is the maximum of $(t - s)^{-1/2}|Y_o(s, t)| - C(t - s)$ over all $(s, t) \in \Pi_o$, this implies that $T_o < \infty$ almost surely. On the other hand, for any integer $\ell \geq 1/u$ the variables $Z_i := \ell^{1/2}Y_o((i - 1)/\ell, i/\ell)$, $1 \leq i \leq \ell$, are independent and standard Gaussian. It is a well-known fact that the maximum of these variables equals $(2 \log \ell)^{1/2} + o_p(1) = C(1/\ell) + o_p(1)$ as $\ell \rightarrow \infty$. Since $T_o \geq \max_{i \leq \ell} Z_i - C(1/\ell)$, this entails that $T_o \geq 0$ almost surely.

Now we define the grid $\Gamma_n := \{0, 1/n, 2/n, \dots, 1\}$ and the index set

$$\Pi_n := \{(s, t) : s, t \in \Gamma_n, 0 < t - s \leq u\}.$$

For $(s, t) \in \Pi_n$ let

$$Y_n(s, t) := n^{-1/2} \sum_{i=ns+1}^{nt} \beta\left(\frac{i - ns}{nt - ns + 1}\right) q\left(\frac{R_{ns, nt}^*(i)}{nt - ns + 1}\right).$$

The convergence in distribution of T^* to T_o can be derived from the following two claims:

Claim 2.1 (a). Conditions (7.5) and (7.6) of Proposition 7.1 are satisfied for $(Y, \Pi) = (Y_n, \Pi_n)$ with constants L and M not depending on n . Moreover,

$$(7.8) \quad (t - s)^{-1/2}|Y_n(s, t)| \leq K(n(t - s))^{1/2} \quad \text{for all } n \geq 2, (s, t) \in \Pi_n.$$

Throughout this proof K denotes a generic positive constant depending only on β, q and $(h_d)_{d=1}^\infty$. Its value may be different in different expressions.

Claim 2.1 (b). The process Y_n , suitably extended to a process on Π_o , converges in distribution to the centered Gaussian process Y_o defined in (7.7).

Before proving these two claims let us derive the assertion of Theorem 2.1. Note that

$$T^* = \max_{(s,t) \in \Pi_n} \left(\frac{(1 + h_n(t-s))|Y_{ns,nt}|}{(t-s)^{1/2}} - C(t-s) \right).$$

By our assumption that $h_d = O(d^{-1})$ and (7.8),

$$\begin{aligned} & \max_{(s,t) \in \Pi_n : t-s \leq \ell/n} \left(\frac{(1 + h_n(t-s))|Y_n(s,t)|}{(t-s)^{1/2}} - C(t-s) \right) \\ & \leq K\ell^{1/2} - (2\log n - 2\log \ell)^{1/2} \rightarrow -\infty \quad \text{for any fixed integer } \ell > 0, \\ & \max_{(s,t) \in \Pi_n : t-s \geq \ell(n)/n} \frac{|h_n(t-s)||Y_n(s,t)|}{(t-s)^{1/2}} \\ & \leq K\ell(n)^{-1/2} \rightarrow 0 \quad \text{whenever } \ell(n) \rightarrow \infty. \end{aligned}$$

Consequently it suffices to consider the case $h_d = 0$, i.e.

$$T^* = \max_{(s,t) \in \Pi_n} \left((t-s)^{-1/2}|Y_n(s,t)| - C(t-s) \right).$$

Let the latter maximum be attained at a random index $(\hat{s}_n, \hat{t}_n) \in \Pi_n$. Since the additional factor $D(t-s) = (\log(e/(t-s)))^{-1/2} \log \log(e^e/(t-s))$ in Proposition 7.1 converges to zero as $t-s \downarrow 0$, it follows from Claim 2.1 (a) and Proposition 7.1 that for any fixed number $r > 0$,

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P} \left\{ T^* \geq r, \hat{t}_n - \hat{s}_n \leq \delta \right\} = 0.$$

Thus one can conclude from Claim 2.1 (b) and the Continuous Mapping Theorem that

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \mathbb{P} \{ T^* \geq r \} \\ & \leq \limsup_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P} \left\{ T^* \geq r, \hat{t}_n - \hat{s}_n \geq \delta \right\} \\ & \leq \limsup_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P} \left\{ \max_{(s,t) \in \Pi_n : t-s \geq \delta} \left((t-s)^{-1/2}|Y_n(s,t)| - C(t-s) \right) \geq r \right\} \\ & \leq \limsup_{\delta \downarrow 0} \mathbb{P} \left\{ \sup_{(s,t) \in \Pi_o : t-s \geq \delta} \left((t-s)^{-1/2}|Y_o(s,t)| - C(t-s) \right) \geq r \right\} \\ & \leq \mathbb{P} \{ T_o \geq r \}. \end{aligned}$$

Moreover, for arbitrary $\tilde{r} \in \mathbb{R}$,

$$\begin{aligned}
& \limsup_{n \rightarrow \infty} \mathbb{P}\{T^* \leq \tilde{r}\} \\
& \leq \inf_{\delta > 0} \limsup_{n \rightarrow \infty} \mathbb{P}\left\{\max_{(s,t) \in \Pi_n : t-s \geq \delta} \left((t-s)^{-1/2}|Y_n(s,t)| - C(t-s)\right) \leq \tilde{r}\right\} \\
& \leq \inf_{\delta > 0} \mathbb{P}\left\{\sup_{(s,t) \in \Pi_o : t-s \geq \delta} \left((t-s)^{-1/2}|Y_o(s,t)| - C(t-s)\right) \leq \tilde{r}\right\} \\
& = \mathbb{P}\{T_o \leq \tilde{r}\}.
\end{aligned}$$

In particular, $\mathbb{P}\{T^* \leq \tilde{r}\} \rightarrow 0$ for any $\tilde{r} < 0$. These conclusions entail that T^* converges in distribution to T_o . \square

Proof of Claim 2.1 (a). According to Corollary 6.6, for any $(s, t) \in \Pi_n$ and $\eta > 0$,

$$(7.9) \quad \mathbb{P}\{|Y_n(s, t)| \geq (t-s)^{1/2}\eta\} \leq 2 \exp\left(-\frac{\eta^2 \Sigma_{n(t-s)}(\beta)^2}{2n(t-s)}\right).$$

Here we utilize the fact that $\{1, 2, \dots, n(t-s)\} \ni j \mapsto R_{ns, nt}^*(ns+j)$ is uniformly distributed on the set $\mathcal{S}_{n(t-s)}$. Moreover, the Cauchy-Schwarz inequality and Proposition 6.5 yield the inequality

$$(7.10) \quad |Y_n(s, t)| \leq (t-s)^{1/2} \Sigma_{n(t-s)}(\beta).$$

Now, for any function g on \mathbb{R} with finite total variation $\text{TV}(g)$,

$$(7.11) \quad \left|d^{-1} \sum_{j=1}^d g(j/(d+1)) - \int_0^1 g(x) dx\right| \leq \text{TV}(g)/d,$$

a well-known inequality from calculus. Since $\text{TV}(\mathcal{P}) < \infty$, this entails that

$$(7.12) \quad \left|d^{-1} \Sigma_d(\beta)^2 - 1\right| \leq K/d.$$

Now (7.8) is a consequence of (7.10) and (7.12). The fact that Condition (7.5) holds uniformly in n follows from (7.9), (7.10) and (7.12):

$$\begin{aligned}
\mathbb{P}\{|Y_n(s, t)| \geq (t-s)^{1/2}\eta\} & \leq 1\{\eta^2 \leq Kn(t-s)\} 2 \exp\left(-\left(1 - \frac{K}{n(t-s)}\right) \frac{\eta^2}{2}\right) \\
& \leq 2 \exp\left(-\frac{\eta^2}{2} + K\right) \\
& = K \exp\left(-\frac{\eta^2}{2}\right).
\end{aligned}$$

In order to verify Condition (7.6) it suffices to consider pairs (s, t) and (s', t') such that $s = s'$ or $t = t'$. For note that

$$\begin{aligned}
|Y_n(s, t) - Y_n(s', t')| & \leq |Y_n(s, t) - Y_n(s', t)| + |Y_n(s', t) - Y_n(s', t')|, \\
\rho((s, t), (s', t'))^2 & = \rho((s, t), (s', t))^2 + \rho((s', t), (s', t'))^2,
\end{aligned}$$

whence $\mathbb{P}\{|Y_n(s, t) - Y_n(s', t')| > \rho((s, t), (s', t'))\eta\}$ is bounded by

$$\begin{aligned} & \mathbb{P}\{|Y_n(s, t) - Y_n(s', t)| > \rho((s, t), (s', t))\eta/2\} \\ & + \mathbb{P}\{|Y_n(s', t) - Y_n(s', t')| > \rho((s', t), (s', t'))\eta/2\}. \end{aligned}$$

Thus let $s \leq s' < t' \leq t$ with $s = s'$ or $t = t'$. Because of the special form of our processes Y_n it even suffices to verify (7.6) in the special case

$$s = s' = 0 < t' < t = 1 \quad \text{with} \quad t' = d/n \in]0, 1[,$$

provided that the constant M we come up with does not depend on n . This can be verified by symmetry considerations and a simple rescaling argument. Thus we have to show that

$$\mathbb{P}\{|Z - Z'| > \Delta\eta\} \leq K \exp(-\eta^2/K) \quad \text{for all } \eta \geq 0,$$

where

$$\begin{aligned} Z := Y_n(0, 1) &= \sum_{i=1}^n b_i q\left(\frac{R^*(i)}{n+1}\right) \quad \text{with } b_i := n^{-1/2} \beta\left(\frac{i}{n+1}\right), \\ Z' := Y_n(0, t') &= \sum_{i=1}^n b'_i q\left(\frac{R_{0d}^*(i)}{d+1}\right) \quad \text{with } b'_i := n^{-1/2} \beta\left(\frac{i}{d+1}\right), \\ \Delta &:= (1 - d/n)^{1/2}. \end{aligned}$$

In order to prove such a tail inequality for $|Z - Z'|$ we introduce the additional random variable

$$Z'' := \sum_{i=1}^n b'_i q\left(\frac{R^*(i)}{n+1}\right)$$

and verify that for all $\eta \geq 0$,

$$(7.13) \quad \mathbb{P}\{|Z - Z''| > \Delta\eta\} \leq K \exp(-\eta^2/K),$$

$$(7.14) \quad \mathbb{P}\{|Z'' - Z'| > \Delta\eta\} \leq K \exp(-\eta^2/K).$$

It follows from Corollary 6.6 that for arbitrary $\eta > 0$,

$$\mathbb{P}\{|Z - Z''| > \eta\} \leq K \exp\left(-\frac{\eta^2}{K\|b - b'\|_2^2}\right).$$

Thus (7.13) would follow from

$$(7.15) \quad \|b - b'\|_2^2 \leq K\Delta^2.$$

One can write

$$\begin{aligned} \|b - b'\|_2^2 &= \frac{1}{n} \sum_{i=1}^d \left(\beta\left(\frac{i}{n+1}\right) - \beta\left(\frac{i}{d+1}\right)\right)^2 + \frac{1}{n} \sum_{i=d+1}^n \beta\left(\frac{i}{n+1}\right)^2 \\ &\leq \frac{1}{n} \sum_{i=1}^d \left(\beta\left(\frac{i}{n+1}\right) - \beta\left(\frac{i}{d+1}\right)\right)^2 + K\Delta^2. \end{aligned}$$

In view of Remark 1 on Theorem 2.1 we utilize only finiteness of $\text{TV}(\beta)$. Let $\bar{\mu}$ be a finite measure on $[0, 1]$ such that $|\beta(b) - \beta(a)| \leq \bar{\mu}([a, b])$ for $0 \leq a < b \leq 1$. Then, by (7.11) and Fubini's theorem,

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^d \left(\beta\left(\frac{i}{n+1}\right) - \beta\left(\frac{i}{d+1}\right) \right)^2 \\
& \leq \frac{K}{n} + \frac{d}{n} \int_0^1 \left(\beta\left(\frac{d+1}{n+1}x\right) - \beta(x) \right)^2 dx \\
& \leq K\Delta^2 + \frac{d}{n} \int_0^1 \bar{\mu}(J(x))^2 dx \\
& \leq K\Delta^2 + \frac{d}{n} \int \text{Leb}\{x \in [0, 1] : y \in J(x), z \in J(x)\} \bar{\mu}(dy) \bar{\mu}(dz) \\
& \leq K\Delta^2 + K \frac{d}{n} \sup_{y \in [0, 1]} \text{Leb}\{x \in [0, 1] : y \in J(x)\},
\end{aligned}$$

where $J(x) := \left[(d+1)x/(n+1), x \right]$. But for any $y \in [0, 1]$, the Lebesgue measure of the set $\{x \in [0, 1] : y \in J(x)\} = \left[y, (n+1)y/(d+1) \right]$ equals $(n-d)y/(d+1) \leq n\Delta^2/d$. This completes our proof of (7.15).

As for assertion (7.14) about $Z'' - Z'$, we write $R^*(i) = R_{0d}^*(i) + \tilde{R}_{0d}(i)$, where

$$\tilde{R}_{0d}(i) := \sum_{h>d} 1\{U_h \leq U_i\}.$$

Then with $b_i'' := b_i' - d^{-1} \sum_{h=1}^d b_h'$ one may write

$$Z'' - Z' = \sum_{i=1}^d b_i'' \left(q\left(\frac{R_{0d}^*(i) + \tilde{R}_{0d}(i)}{n+1}\right) - q\left(\frac{R_{0d}^*(i)}{d+1}\right) \right).$$

Now Proposition 6.3 is applicable to the conditional distribution of $(R_{0d}^*(i))_{i \leq d}$ given $(\tilde{R}_{0d}(i))_{i \leq d}$, which is uniform on the set \mathcal{S}_d . One easily verifies that $|b_i''| \leq Kn^{-1/2}$ for $1 \leq i \leq n$. Thus one can apply Proposition 6.3 as follows:

$$\begin{aligned}
& \mathbb{P}\{|Z'' - Z'| > \Delta\eta\} \\
& \leq \mathbb{E} K \exp \left(-K^{-1} \Delta^2 \eta^2 n / \sum_{j=1}^d \left(q\left(\frac{\tilde{R}_{0d}(j) + j}{n+1}\right) - q\left(\frac{j}{d+1}\right) \right)^2 \right) \\
& \leq K \exp \left(-K^{-1} \Delta^2 \eta^2 n / \sum_{j \leq n-n\Delta^2} \left(q\left(\frac{j+n\Delta^2}{n+1}\right) - q\left(\frac{j}{n+1}\right) \right)^2 \right) \\
& \leq K \exp(-\eta^2/K),
\end{aligned}$$

which is (7.14). The latter displayed inequality is a consequence of (2.4). For $(q(x+\Delta^2) - q(x))^2$

is nondecreasing in $x \in [1/2 - \Delta^2/2, 1[$ and nonincreasing in $x \in]0, 1/2 - \Delta^2/2]$, whence

$$\begin{aligned} & \left(q\left(\frac{j+n\Delta^2}{n+1}\right) - q\left(\frac{j}{n+1}\right) \right)^2 \\ & \leq \begin{cases} (n+1) \int_{j/(n+1)}^{(j+1)/(n+1)} (q(x+\Delta^2) - q(x))^2 dx & \text{if } \frac{j}{n+1} \geq 1/2 - \Delta^2/2, \\ (n+1) \int_{(j-1)/(n+1)}^{j/(n+1)} (q(x+\Delta^2) - q(x))^2 dx & \text{if } \frac{j}{n+1} \leq 1/2 - \Delta^2/2. \end{cases} \end{aligned}$$

Thus

$$\frac{1}{n} \sum_{j \leq n - n\Delta^2} \left(q\left(\frac{j+n\Delta^2}{n+1}\right) - q\left(\frac{j}{n+1}\right) \right)^2 \leq 2 \int_0^{1-\Delta^2} (q(x+\Delta^2) - q(x))^2 dx \leq K\Delta^2. \quad \square$$

Proof of Claim 2.1 (b). It follows from Claim 2.1 (a) and standard chaining inequalities that the sequence $(Y_n)_n$ is stochastically equicontinuous in the sense that

$$\sup_{n \geq 2} \mathbb{E} \sup_{a, b \in \Pi_n : \rho(a, b) < \delta} |Y_n(a) - Y_n(b)| \rightarrow 0 \quad \text{as } \delta \downarrow 0;$$

cf. van der Vaart and Wellner (1996, Chapter 2.2). Hence it suffices to prove weak convergence of the finite-dimensional distributions of $(Y_n)_n$. Precisely, for any integer $d \geq 1$ and points $(s_{ni}, t_{ni}) \in \Pi_n$, $1 \leq i \leq d$, we have to show that

$$(7.16) \quad (Y_n(s_{ni}, t_{ni}))_{i=1}^d \rightarrow_{\mathcal{L}} (Y_o(s_i, t_i))_{i=1}^d$$

as $n \rightarrow \infty$ and $(s_{ni}, t_{ni}) \rightarrow (s_i, t_i) \in \Pi_o$ for $1 \leq i \leq d$.

Since β is bounded one can apply Proposition 6.4 with $n(t-s)$ in place of n in order to show that $\mathbb{E}(Y_n(s, t) - Y'_n(s, t))^2 \leq h(n(t-s))$, where $\lim_{d \rightarrow \infty} h(d) = 0$ and

$$Y'_n(s, t) := n^{-1/2} \sum_{i=1}^n \beta\left(\frac{i - ns}{n(t-s) + 1}\right) q\left(\frac{\lceil n(t-s)U_i \rceil}{n(t-s) + 1}\right).$$

Finally, if

$$Y''_n(s, t) := n^{-1/2} \sum_{i=1}^n \beta\left(\frac{i - ns}{n(t-s) + 1}\right) q(U_i),$$

then

$$\mathbb{E}(Y'_n(s, t) - Y''_n(s, t))^2 \leq \sup(\beta^2) h'(n(t-s)),$$

where

$$h'(d) := \mathbb{E}\left(q\left(\frac{\lceil dU_1 \rceil}{d+1}\right) - q(U_1)\right)^2 \rightarrow 0 \quad \text{as } d \rightarrow \infty.$$

The latter claim follows from Scheffé's theorem and the following two facts:

$$\begin{aligned} \mathbb{E} q\left(\frac{\lceil dU_1 \rceil}{d+1}\right)^2 &= \text{Var}\left(q\left(\frac{\lceil dU_1 \rceil}{d+1}\right)\right) \leq \sigma(q)^2 = 1 \quad \text{by Corollary 6.6,} \\ \lim_{d \rightarrow \infty} q\left(\frac{\lceil dU_1 \rceil}{d+1}\right) &= q(U_1) \quad \text{by continuity of } q. \end{aligned}$$

Thus one may replace Y_n in (7.16) with Y_n'' . But then the assertion is a simple consequence of the multivariate version of Lindeberg's Central Limit Theorem. \square

Proof of Theorem 3.1. By symmetry it suffices to consider the case $\lambda < 0$. Let $\delta_n := (\log(n)/n)^{1/(2\gamma+1)}$. We consider two special intervals, $J_d = [x_{j_d+1}, x_{k_d}]$ for $d = 1, 2$, where

$$\begin{aligned} j_1 &:= \min\{i : x_i \geq x_o - D\delta_n\} \quad \text{and} \quad k_1 := \max\{i : x_i \leq x_o\}, \\ j_2 &:= \min\{i : x_i \geq x_o\} \quad \text{and} \quad k_2 := \max\{i : x_i \leq x_o + D\delta_n\} \end{aligned}$$

with some constant $D > 0$ to be specified later. Then

$$m_d := k_d - j_d = Dn\delta_n(1 + o(1)) \quad \text{and} \quad \max\{|x_i - x_o| : x_i \in J_d\} = O(\delta_n).$$

Thus it follows from Condition (3.2) and continuity of G that

$$\max\{\|G_i - G_j\|_\infty : x_i, x_j \in J_1 \cup J_2\} = o(1).$$

Hence Proposition 6.7 implies that

$$\text{Var}(T_{j_d, k_d}) \leq \Sigma_{m_d}(\beta)^2(1 + o(1)) = (D\sigma(\beta)^2 + o(1))n\delta_n.$$

On the other hand, Condition (3.1) implies that

$$\begin{aligned} \mathbb{E} T_{j_2, k_2} &= \frac{1}{m_2 + 1} \sum_{j_2 < a < b \leq k_2} \left(\beta\left(\frac{b - j_2}{m_2 + 1}\right) - \beta\left(\frac{a - j_2}{m_2 + 1}\right) \right) \left(2 \int G_a dG_b - 1 \right) \\ &= \frac{1}{m_2 + 1} \sum_{j_2 < a < b \leq k_2} \left(\beta\left(\frac{b - j_2}{m_2 + 1}\right) - \beta\left(\frac{a - j_2}{m_2 + 1}\right) \right) \\ &\quad \cdot \left(2 \int G\left(x + f\left(\frac{b}{n}\right) - f\left(\frac{a}{n}\right)\right) G(dx) - 1 \right) \\ &\geq \frac{g + o(1)}{m_2 + 1} \sum_{j_2 < a < b \leq k_2} \left(\beta\left(\frac{b - j_2}{m_2 + 1}\right) - \beta\left(\frac{a - j_2}{m_2 + 1}\right) \right) \left(f\left(\frac{b}{n}\right) - f\left(\frac{a}{n}\right) \right) \\ &= \frac{g|\lambda| + o(1)}{m_2 + 1} \sum_{j_2 < a < b \leq k_2} \left(\beta\left(\frac{b - j_2}{m_2 + 1}\right) - \beta\left(\frac{a - j_2}{m_2 + 1}\right) \right) \left(\left(\frac{b}{n} - x_o\right)^\gamma - \left(\frac{a}{n} - x_o\right)^\gamma \right) \\ &\quad + o(m_2\delta_n^\gamma) \\ &= \frac{g|\lambda| + o(1)}{m_2 + 1} \sum_{0 < a < b \leq m_2} \left(\beta\left(\frac{b}{m_2 + 1}\right) - \beta\left(\frac{a}{m_2 + 1}\right) \right) \left(\left(\frac{b}{n}\right)^\gamma - \left(\frac{a}{n}\right)^\gamma \right) \\ &\quad + o(m_2\delta_n^\gamma) \\ &= \frac{D^\gamma g|\lambda| + o(1)}{m_2 + 1} \sum_{0 < a < b \leq m_2} \left(\beta\left(\frac{b}{m_2 + 1}\right) - \beta\left(\frac{a}{m_2 + 1}\right) \right) \\ &\quad \cdot \left(\left(\frac{b}{m_2 + 1}\right)^\gamma - \left(\frac{a}{m_2 + 1}\right)^\gamma \right) \delta_n^\gamma + o(m_2\delta_n^\gamma) \\ &= (D^\gamma g|\lambda|h + o(1))m_2\delta_n^\gamma + o(m_2\delta_n^\gamma) \\ &= (D^{1+\gamma}g|\lambda|h + o(1))n\delta_n^{1+\gamma}, \end{aligned}$$

where g denotes the limes inferior in (3.1), and

$$h := \int_{\{(u,v): 0 < u < v < 1\}} (\beta(v) - \beta(u))(v^\gamma - u^\gamma) du dv > 0.$$

Analogously one can show that

$$\mathbb{E} T_{j_1, k_1} \leq -n \delta_n^{1+\gamma} (D^{1+\gamma} g |\lambda| h + o(1)).$$

Hence

$$\begin{aligned} & \mathbb{P} \left\{ \inf \left\{ y - x : x_o \in [x, y] \in \mathcal{M}^-(\alpha) \right\} > 2D\delta_n \right\} \\ & \leq \mathbb{P} \left\{ [x_{j_1+1}, x_{k_2}] \notin \mathcal{M}^-(\alpha) \right\} \\ & \leq \mathbb{P} \left\{ s_{m_1} T_{j_1 k_1} \geq -(c_{m_1} + \kappa(\alpha)) \text{ or } s_{m_2} T_{j_2 k_2} \leq c_{m_2} + \kappa(\alpha) \right\} \\ & \leq \sum_{\ell=1,2} \mathbb{P} \left\{ s_{m_\ell} |T_{j_\ell k_\ell} - \mathbb{E} T_{j_\ell k_\ell}| \geq \left(s_{m_\ell} |\mathbb{E} T_{j_\ell k_\ell}| - c_{m_\ell} - \kappa(\alpha) \right)_+ \right\} \\ & \leq \sum_{\ell=1,2} s_{m_\ell}^2 \text{Var}(T_{j_\ell k_\ell}) / \left(s_{m_\ell} |\mathbb{E} T_{j_\ell k_\ell}| - c_{m_\ell} - \kappa(\alpha) \right)_+^2 \\ & \leq 2(\log n)^{-1} / \left(D^{1/2+\gamma} g |\lambda| h - 2^{1/2} \sigma(q) + o(1) \right)_+^2 \\ & = o(1) \quad \text{provided that } D^{1/2+\gamma} g |\lambda| h > 2^{1/2} \sigma(q). \quad \square \end{aligned}$$

Proof of Proposition 6.5. Let $U_{(1)} < U_{(2)} < \dots < U_{(n)}$ be the order statistics of U , so that $U_i = U_{(R^*(i))}$. Then it follows from Jensen's inequality that

$$\sigma(q)^2 = \mathbb{E} q(U_1)^2 = \mathbb{E} \mathbb{E} \left(q(U_1)^2 \mid R^* \right) \geq \mathbb{E} \mathbb{E} (q(U_1) \mid R^*)^2 = n^{-1} \sum_{j=1}^n Q_j^2,$$

where

$$Q_j := \mathbb{E} q(U_{(j)}).$$

Moreover

$$\begin{aligned} \mathbb{E} \exp \left(\sum_{i=1}^n b_i q(U_i) \right) &= \mathbb{E} \mathbb{E} \left(\exp \left(\sum_{i=1}^n b_i q(U_i) \right) \mid R^* \right) \\ &\geq \mathbb{E} \exp \left(\sum_{i=1}^n b_i \mathbb{E} (q(U_i) \mid R^*) \right) \\ &= \mathbb{E} \exp \left(\sum_{i=1}^n b_i Q_{R^*(i)} \right), \end{aligned}$$

Now we consider the conditional distribution of R^* given the two-point sets

$$M_j := \{R^{*-1}(j), R^{*-1}(n+1-j)\}, \quad 1 \leq j \leq m := \lfloor n/2 \rfloor.$$

Let $M_j = \{M'_j, M''_j\}$ with $M'_j < M''_j$. Conditional on $M := (M_j)_{1 \leq j \leq m}$, the pairs

$$(R^{*-1}(j), R^{*-1}(n+1-j)), \quad 1 \leq j \leq m,$$

are stochastically independent with

$$\begin{aligned} & \mathbb{P}\left((R^{*-1}(j), R^{*-1}(n+1-j)) = (M'_j, M''_j) \mid M\right) \\ &= \mathbb{P}\left((R^{*-1}(j), R^{*-1}(n+1-j)) = (M''_j, M'_j) \mid M\right) \\ &= 1/2. \end{aligned}$$

Since $Q_{n+1-j} = -Q_j$,

$$\begin{aligned} \mathbb{E} \exp\left(\sum_{i=1}^n b_i Q_{R^*(i)}\right) &= \mathbb{E} \exp\left(\sum_{i=1}^n b_{R^{*-1}(i)} Q_j\right) \\ &= \mathbb{E} \exp\left(\sum_{j=1}^m (b_{R^{*-1}(j)} - b_{R^{*-1}(n+1-j)}) Q_j\right) \\ &= \mathbb{E} \mathbb{E}\left(\exp\left(\sum_{j=1}^m (b_{R^{*-1}(j)} - b_{R^{*-1}(n+1-j)}) Q_j\right) \mid M\right) \\ &= \mathbb{E} \prod_{j=1}^m \cosh\left((b_{M'_j} - b_{M''_j}) Q_j\right). \end{aligned}$$

Analogously, $q((n+1-j)/(n+1)) = -q(j/(n+1))$, whence

$$\mathbb{E} \exp\left(\sum_{i=1}^n b_i q\left(\frac{R^*(i)}{n+1}\right)\right) = \mathbb{E} \prod_{j=1}^m \cosh\left((b_{M'_j} - b_{M''_j}) q\left(\frac{j}{n+1}\right)\right).$$

Since $q \geq 0$ on $[1/2, 1[$ the proof of Proposition 6.5 is complete when we have shown that

$$(7.17) \quad q\left(\frac{j}{n+1}\right) \leq Q_j \quad \text{for } (n+1)/2 < j \leq n.$$

In order to prove (7.17) note that the distribution of $U_{(j)}$ is a Beta-distribution with parameters j and $n+1-j$. It has a unimodal density $f_j(x) := B_j x^{j-1} (1-x)^{n-j}$ on $]0, 1[$. Thus it can be represented as a mixture of uniform distributions on intervals $I_j(r) := \{f_j > r\} \subset]0, 1[$:

$$f_j(x) = \int_0^{\max(f_j)} \text{Leb}(I_j(r)) \frac{1_{\{x \in I_j(r)\}}}{\text{Leb}(I_j(r))} dr,$$

where $\text{Leb}(\cdot)$ stands for Lebesgue measure. Elementary considerations show that the midpoint $m_j(r)$ of $I_j(r)$ is not smaller than $1/2$. Therefore, since $\mathbb{E} U_{(j)} = j/(n+1)$,

$$\begin{aligned} q\left(\frac{j}{n+1}\right) &= q\left(\int_0^1 x f_j(x) dx\right) \\ &= q\left(\int_0^{\max(f_j)} \text{Leb}(I_j(r)) m_j(r) dr\right) \\ &\leq \int_0^{\max(f_j)} \text{Leb}(I_j(r)) q(m_j(r)) dr \end{aligned}$$

by Jensen's inequality. For fixed $r \in [0, \max(f_j)[$ let $I_j(r) =]u - \delta, u + \delta[$ with $u = m_j(r) \in [1/2, 1[$ and $0 < \delta \leq 1 - u$. Then

$$\begin{aligned} \int_{I_j(r)} q(x) dx - \text{Leb}(I_j(r))q(m_j(r)) &= \int_{-\delta}^{\delta} (q(u+x) - q(u)) dx \\ &= \int_0^{\delta} \int_0^x (q'(u+s) - q'(u-s)) ds dx \\ &\geq 0, \end{aligned}$$

where $q' \geq 0$ denotes the L^1 -derivative of q on $]0, 1[$. For without loss of generality let q be non-decreasing on $[1/2, 1[$ and symmetric around $1/2$. Then elementary considerations show that $q'(u+s) - q'(u-s) \geq 0$ for $1/2 \leq u < 1$ and $0 < s < 1 - u$. Consequently,

$$q\left(\frac{j}{n+1}\right) \leq \int_0^{\max(f_j)} \int_{I_j(r)} q(x) dx dr = \int_0^1 f_j(x) q(x) dx = Q_j. \quad \square$$

Proof of Corollary 6.6. Let $Z := \sum_{i=1}^n b_i q(R^*(i)/(n+1))$. For $\lambda \in \mathbb{R}$ it follows from Proposition 6.5 applied to λb in place of b that $\mathbb{E} \exp(\lambda Z)$ is not greater than

$$\mathbb{E} \exp\left(\sum_{i=1}^n \lambda b_i q(U_i)\right) = \prod_{i=1}^n M(q, \lambda b_i) \leq \exp(\lambda^2 \sigma(q)^2 \|b\|_2^2 / 2).$$

Thus

$$\begin{aligned} \text{Var}(Z) &= 2 \lim_{\lambda \rightarrow 0} \lambda^{-2} (\mathbb{E} \cosh(\lambda Z) - 1) \\ &\leq 2 \lim_{\lambda \rightarrow 0} \lambda^{-2} (\exp(\lambda^2 \sigma(q)^2 \|b\|_2^2 / 2) - 1) \\ &= \sigma(q)^2 \|b\|_2^2, \end{aligned}$$

and Tshebyshev's inequality yields

$$\begin{aligned} \mathbb{P}\left\{\pm \sum_{i=1}^n b_i q\left(\frac{R^*(i)}{n+1}\right) \geq \eta\right\} &\leq \min_{\lambda \geq 0} \exp(-\lambda \eta) \mathbb{E} \exp\left(\pm \lambda \sum_{i=1}^n b_i q\left(\frac{R^*(i)}{n+1}\right)\right) \\ &\leq \min_{\lambda \geq 0} \exp\left(\lambda^2 \sigma(q)^2 \|b\|_2^2 / 2 - \lambda \eta\right) \\ &= \exp\left(-\frac{\eta^2}{2\sigma(q)^2 \|b\|_2^2}\right). \quad \square \end{aligned}$$

Proof of Proposition 6.7. Elementary considerations show that

$$2 \frac{R(i)}{n+1} - 1 = \frac{1}{n+1} \sum_{j=1}^n \text{sign}(Y_i - Y_j)$$

almost surely. Then

$$\sum_{i=1}^n b_i \left(2 \frac{R(i)}{n+1} - 1\right) = \sum_{i,j} b_i \text{sign}(Y_i - Y_j) = \sum_{i < j} (b_j - b_i) \text{sign}(Y_j - Y_i),$$

and $\mathbb{E} \text{sign}(Y_j - Y_i) = 2\mathbb{P}\{Y_i \leq Y_j\} - 1 = 2 \int G_i dG_j - 1$. It remains to be shown that the variance of the latter sum is not greater than $\|b\|^2(1 + 12\Delta^2)$, where Δ stands for the maximum of all numbers $\|G_i - G_j\|_\infty$. For that purpose we utilize Hoeffding's decomposition. With $H_{ij} := (b_j - b_i) \text{sign}(Y_j - Y_i)$ and $\tilde{H}_{ij} := H_{ij} - \mathbb{E}(H_{ij} | Y_i) - \mathbb{E}(H_{ij} | Y_j) + \mathbb{E} H_{ij}$,

$$\begin{aligned} \text{Var}\left(\sum_{i < j} H_{ij}\right) &= \sum_{i < j} \text{Var}(\tilde{H}_{ij}) + \text{Var}\left(\sum_{i < j} \left(\mathbb{E}(H_{ij} | Y_i) + \mathbb{E}(H_{ij} | Y_j)\right)\right) \\ &\leq \sum_{i < j} (b_j - b_i)^2 + \text{Var}\left(\sum_{i < j} \left(\mathbb{E}(H_{ij} | Y_i) + \mathbb{E}(H_{ij} | Y_j)\right)\right) \\ &\leq n\|b\|_2^2 + \text{Var}\left(\sum_{i < j} \left(\mathbb{E}(H_{ij} | Y_i) + \mathbb{E}(H_{ij} | Y_j)\right)\right). \end{aligned}$$

Moreover, $\mathbb{E}(H_{ij} | Y_i) = (b_j - b_i)(1 - 2G_j(Y_i))$ and $\mathbb{E}(H_{ij} | Y_j) = (b_j - b_i)(2G_i(Y_j) - 1)$, whence

$$\begin{aligned} &\sum_{i < j} \left(\mathbb{E}(H_{ij} | Y_i) + \mathbb{E}(H_{ij} | Y_j)\right) \\ &= \sum_k \sum_\ell (b_\ell - b_k)(1 - 2G_\ell(Y_k)) \\ &= \sum_k \sum_\ell (b_\ell - b_k)(1 - 2U_k) - 2 \sum_k \sum_\ell (b_\ell - b_k)(G_\ell - G_k)(Y_k) \\ &=: A - B. \end{aligned}$$

But $\text{Var}(A) \leq n^2\|b\|_2^2/3$, while

$$\begin{aligned} \text{Var}(B) &\leq 4 \sum_k \mathbb{E}\left(\sum_\ell (b_\ell - b_k)(G_\ell - G_k)(Y_k)\right)^2 \\ &\leq 4 \sum_k \sum_\ell (b_\ell - b_k)^2 \sum_\ell \|G_\ell - G_k\|_\infty^2 \\ &\leq 8n^2\|b\|_2^2\Delta^2. \end{aligned}$$

Hence $\text{Var}(A - B)$ is not greater than

$$(1/3)^{-1} \text{Var}(A) + (2/3)^{-1} \text{Var}(B) \leq n^2\|b\|_2^2(1 + 12\Delta^2). \quad \square$$

Acknowledgement.

The author is grateful to Stephan Weis and Steve Marron for their interest and helpful comments.

References

CHAUDHURI, P. AND J.S. MARRON (1998). SiZer for exploration of structures in curves. to appear in *J. Amer. Statist. Assoc.*

- DONOHU, D.L. (1988). One-sided inference about functionals of a density. *Ann. Statist.* **16**, 1390-1420
- DÜMBGEN, L. (1998). Symmetrization and decoupling of combinatorial random elements. *Statist. Prob. Letters* **39**, 355-361
- DÜMBGEN, L. AND V.G. SPOKOINY (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29**, 124–152
- HÁJEK, J. (1961). Some extensions of the Wald-Wolfowitz-Noether theorem. *Ann. Math. Statist.* **32**, 506-523
- KHAS'MINSKII, R.Z. (1979). Lower bounds for the risk of nonparametric estimates of the mode. In *Contributions to Statistics, Jaroslav Hájek Memorial Volume* (J. Jureckova, ed.), 91-97. Academia, Prague
- VAN DER VAART, A.W. AND J.A. WELLNER (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer, New York

Department of Mathematical Statistics and Actuarial Science

University of Berne

Sidlerstrasse 5

CH-3012 Berne

Switzerland

E-mail: duembgen@stat.unibe.ch

URL: <http://www.imsv.unibe.ch>

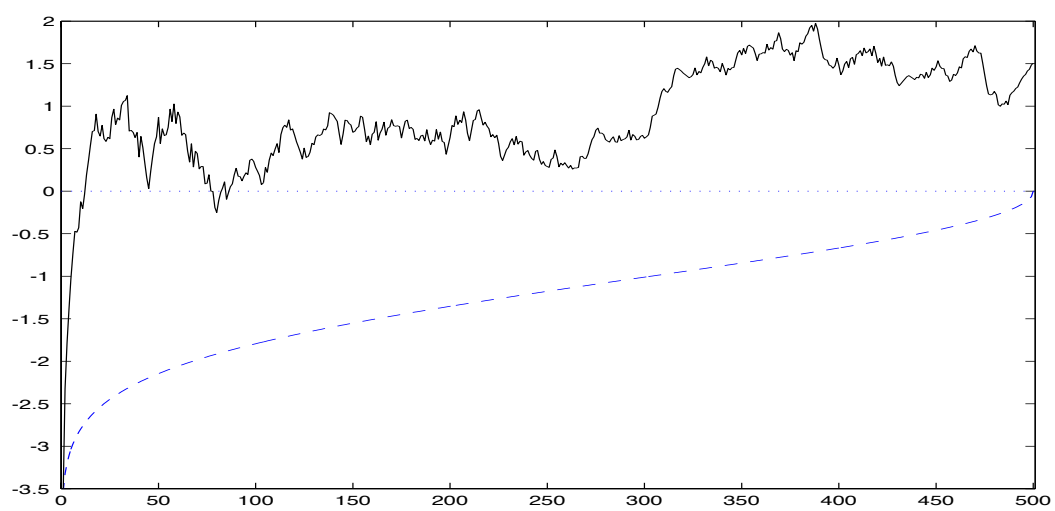
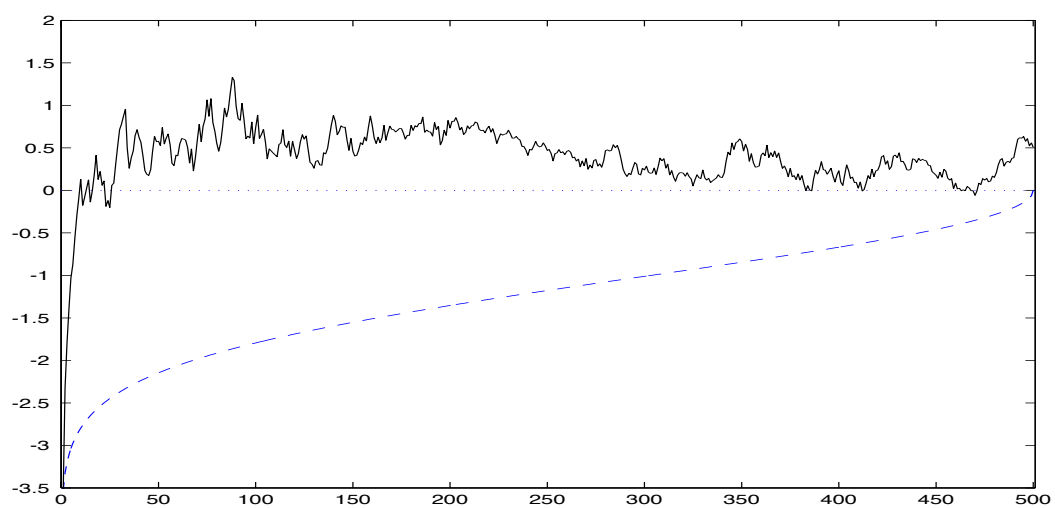
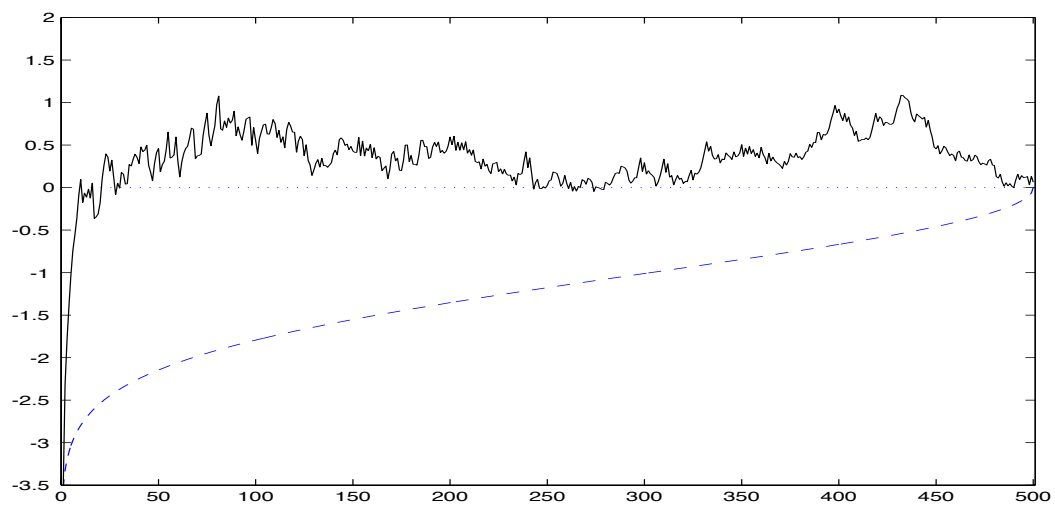


Figure 1. Three realizations of the process $d \mapsto T_d^*$

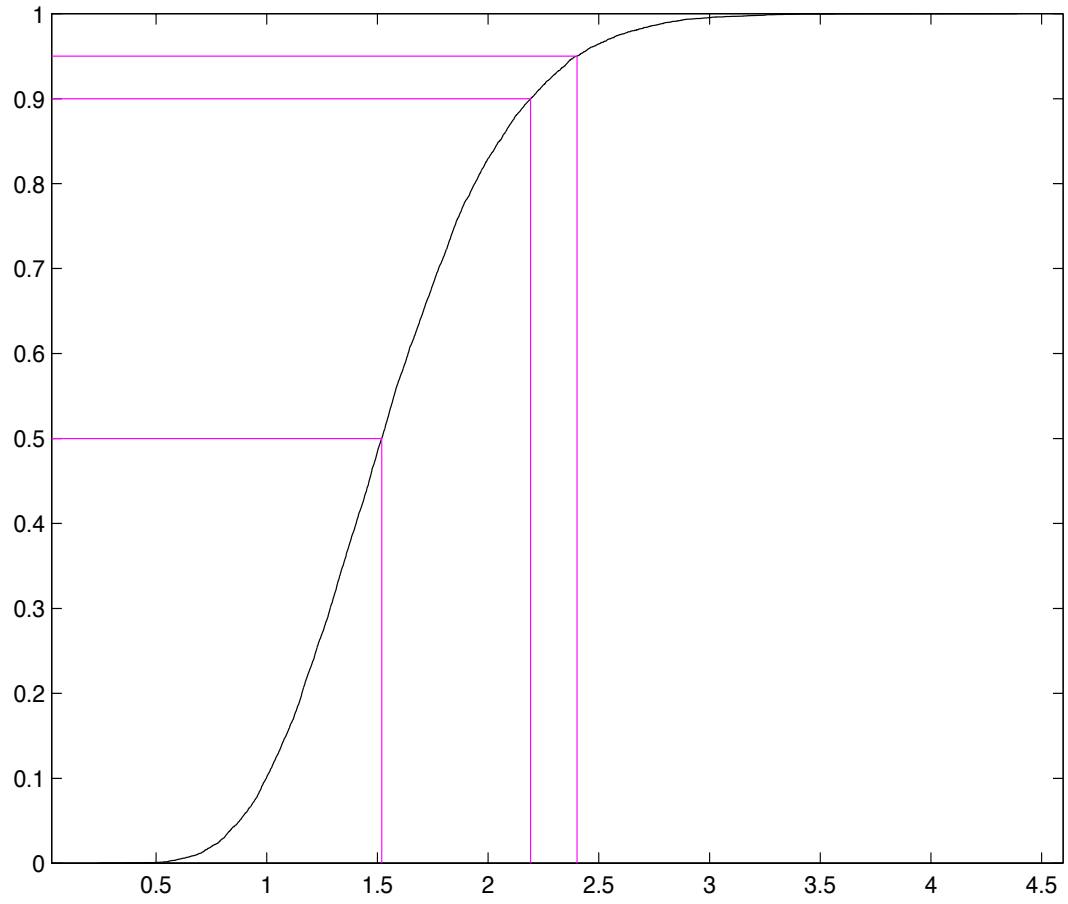


Figure 2. Estimated distribution function of T^* ($n = 500, m = 250$)

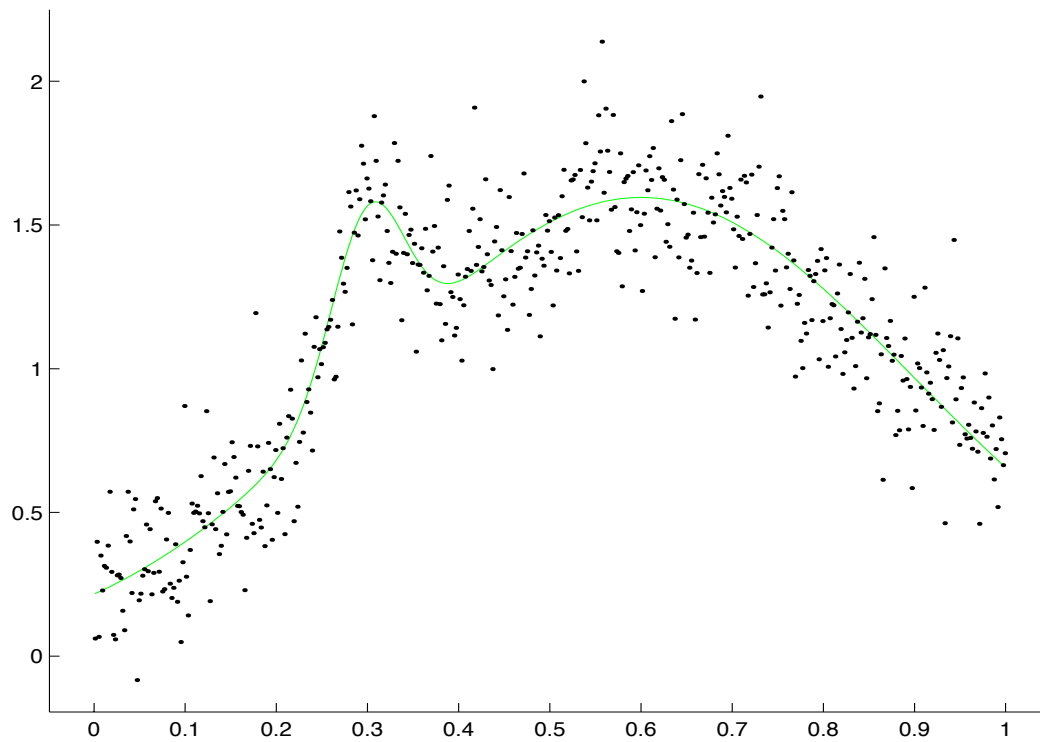


Figure 3. Simulated data (x_i, Y_i) and regression function f ($n = 500$)

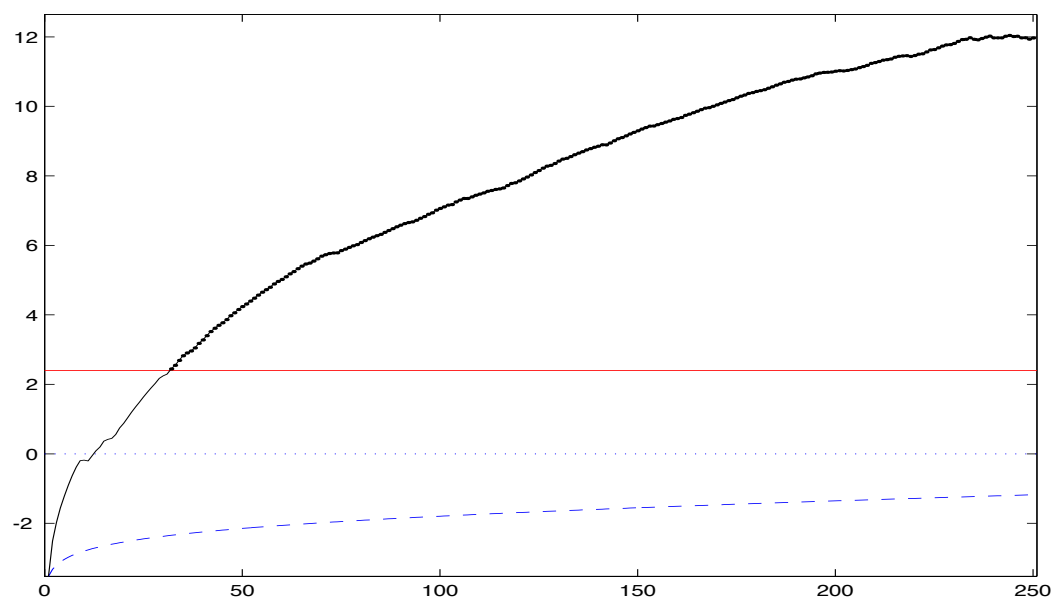


Figure 4. The process $d \mapsto T_d$

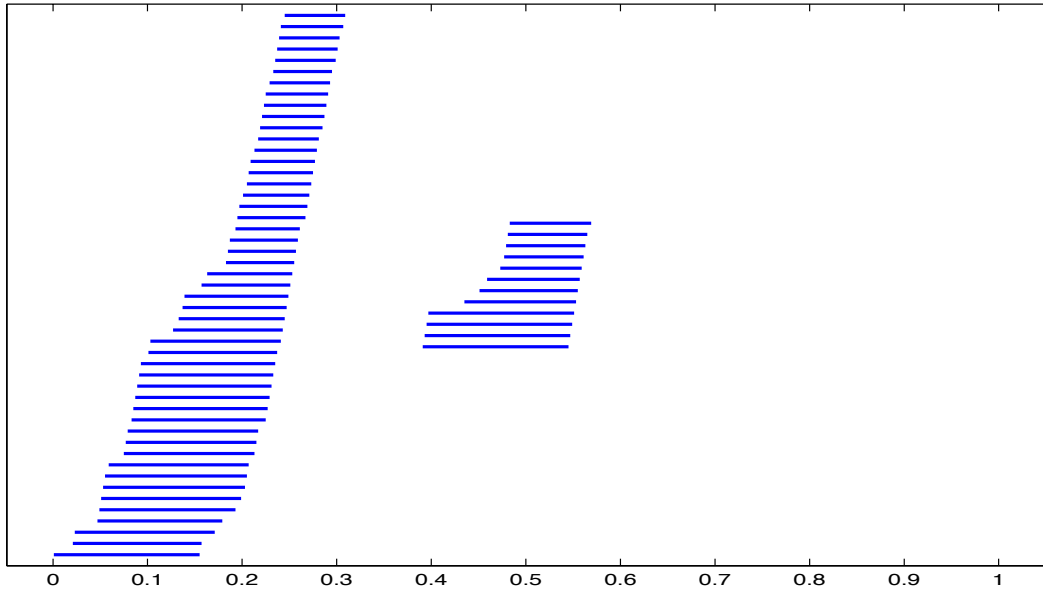


Figure 5a. Minimal intervals in $\mathcal{D}^+(0.05)$

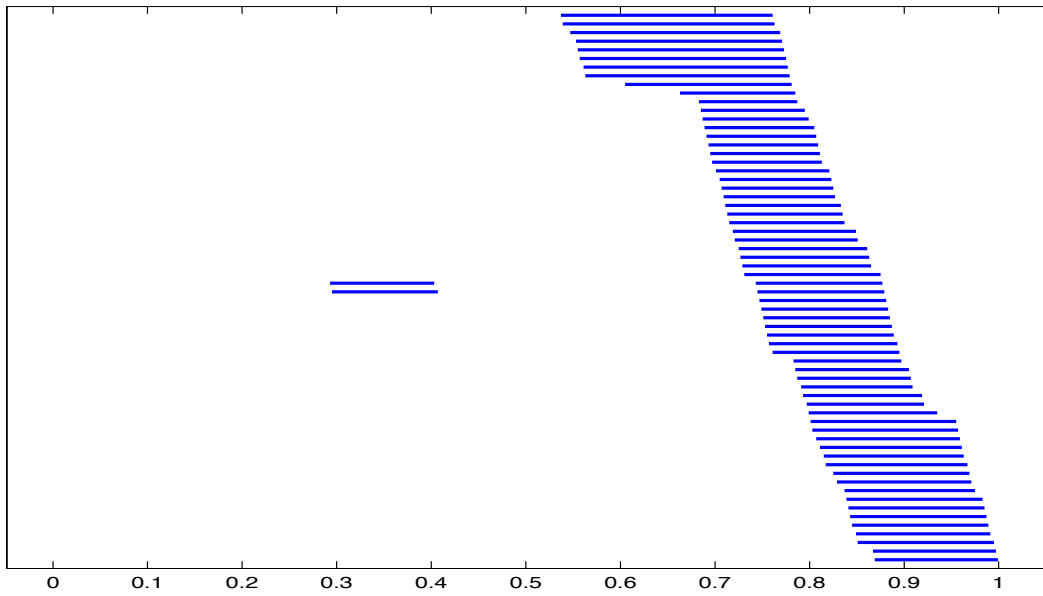


Figure 5b. Minimal intervals in $\mathcal{D}^-(0.05)$

Proposition 6.7. For any $b \in \mathbb{R}^n$,

$$\begin{aligned} \sum_{i=1}^n b_i \left(2 \frac{R(i)}{n+1} - 1 \right) &= \frac{1}{n+1} \sum_{1 \leq i < j \leq n} (b_j - b_i) \text{sign}(Y_j - Y_i) \\ &= \frac{1}{n+1} \sum_{1 \leq i < j \leq n} (b_j - b_i) \left(2 \int G_i dG_j - 1 \right) + Z, \end{aligned}$$

where

$$\mathbb{E} Z^2 \leq \|b\|_2^2 \left(1 + 12 \max_{1 \leq i, j \leq n} \|G_i - G_j\|_\infty^2 \right).$$

Corrected proof of Proposition 6.7. Elementary considerations show that

$$2 \frac{R(i)}{n+1} - 1 = \frac{1}{n+1} \sum_{j=1}^n \text{sign}(Y_i - Y_j)$$

almost surely. Then

$$(n+1) \sum_{i=1}^n b_i \left(2 \frac{R(i)}{n+1} - 1 \right) = \sum_{i,j} b_i \text{sign}(Y_i - Y_j) = \sum_{i < j} (b_j - b_i) \text{sign}(Y_j - Y_i),$$

and $\mathbb{E} \text{sign}(Y_j - Y_i) = 2 \mathbb{P}\{Y_i \leq Y_j\} - 1 = 2 \int G_i dG_j - 1$. It remains to be shown that the variance of the latter sum is not greater than $(n+1)^2 \|b\|^2 (1 + 12\Delta^2)$, where Δ stands for the maximum of all numbers $\|G_i - G_j\|_\infty$. For that purpose we utilize Hoeffding's decomposition. With $H_{ij} := (b_j - b_i) \text{sign}(Y_j - Y_i)$ and $\tilde{H}_{ij} := H_{ij} - \mathbb{E}(H_{ij} | Y_i) - \mathbb{E}(H_{ij} | Y_j) + \mathbb{E} H_{ij}$,

$$\begin{aligned} \text{Var} \left(\sum_{i < j} H_{ij} \right) &= \sum_{i < j} \text{Var}(\tilde{H}_{ij}) + \text{Var} \left(\sum_{i < j} \left(\mathbb{E}(H_{ij} | Y_i) + \mathbb{E}(H_{ij} | Y_j) \right) \right) \\ &\leq \sum_{i < j} (b_j - b_i)^2 + \text{Var} \left(\sum_{i < j} \left(\mathbb{E}(H_{ij} | Y_i) + \mathbb{E}(H_{ij} | Y_j) \right) \right) \\ &\leq n \|b\|_2^2 + \text{Var} \left(\sum_{i < j} \left(\mathbb{E}(H_{ij} | Y_i) + \mathbb{E}(H_{ij} | Y_j) \right) \right). \end{aligned}$$

Moreover, $\mathbb{E}(H_{ij} | Y_i) = (b_j - b_i)(1 - 2G_j(Y_i))$ and $\mathbb{E}(H_{ij} | Y_j) = (b_j - b_i)(2G_i(Y_j) - 1)$, whence

$$\begin{aligned} &\sum_{i < j} \left(\mathbb{E}(H_{ij} | Y_i) + \mathbb{E}(H_{ij} | Y_j) \right) \\ &= \sum_k \sum_\ell (b_\ell - b_k)(1 - 2G_\ell(Y_k)) \\ &= \sum_k \sum_\ell (b_\ell - b_k)(1 - 2U_k) - 2 \sum_k \sum_\ell (b_\ell - b_k)(G_\ell - G_k)(Y_k) \\ &=: A - B. \end{aligned}$$

But $\text{Var}(A) \leq n^2 \|b\|_2^2 / 3$, while

$$\begin{aligned}
\text{Var}(B) &\leq 4 \sum_k \mathbb{E} \left(\sum_\ell (b_\ell - b_k) (G_\ell - G_k) (Y_k) \right)^2 \\
&\leq 4 \sum_k \sum_\ell (b_\ell - b_k)^2 \sum_\ell \|G_\ell - G_k\|_\infty^2 \\
&\leq 8n^2 \|b\|_2^2 \Delta^2.
\end{aligned}$$

Hence $\text{Var}(A - B)$ is not greater than

$$(1/3)^{-1} \text{Var}(A) + (2/3)^{-1} \text{Var}(B) \leq n^2 \|b\|_2^2 (1 + 12\Delta^2).$$

□